DOCUMENT RESUME

ED 480 081 CG 032 654

AUTHOR	Elman, Linda
TITLE	The Future of School Testing: A School District Perspective.
PUB DATE	2003-08-00
NOTE	10p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Accountability; *Educational Assessment; *Educational Testing; *Futures (of Society); School Districts; Technology Uses in Education

ABSTRACT

This chapter presents the author's perspective on the future of educational assessment in her school district. The author reviews issues including how much testing is reasonable (in terms of time and money), how to support more instructionally useful assessment, the impact of technology on assessment, and accountability. (GCP)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 480 081

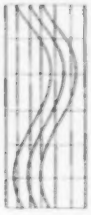
The Future of School Testing: A School District Perspective

By
Linda Elman

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CG032654



Chapter 46
The Future of School Testing
A School District Perspective

Linda Elman

Spring state and district testing is wrapping up now, and the test booklets are being counted, sorted, rubber-banded, and carefully placed in boxes to be sent to a far-off scoring center. Teachers and building administrators are emitting great sighs of relief, glad that the "external" testing cycle is ending for this school year, looking forward to getting results back, but mostly happy to be focusing on classroom-based learning again.

So where is the testing program going from here? The passage of the 2001 NCLB is causing major changes in the district testing calendar, and we try to be patient as the state figures out how we will test all students in grades three through eight in reading and mathematics each year using a standards-based assessment. Once the state assessment program is established we will want to re-examine our district assessment plan to ensure that we are assessing what we deem important, in ways that are time- and cost-efficient, and most importantly, we will want to provide information that improves the learning process for our students.

Five years ago in this school district, teachers and building administrators perceived state and district testing as something that *they* made us do. Now, although the testing is still seen as an external imposition that necessitates modifying school and classroom schedules and procedures, the results are highly valued. Teachers and administrators spend many hours analyzing individual and group results, identifying building and grade-level strengths and weaknesses, and planning changes to improve student achievement.

The dilemma now becomes how much testing is reasonable—in terms of time and money, when it should be conducted, and how it should be conducted. For years this district has been doing direct writing assessment at grades that varied from year to year. Over the last two years the testing migrated to the fall of grades 4, 7, and 10, the same grades that are tested each spring on the state-developed standards-based assessment. The writing assessment is given in the first month of

school, with scored papers returned about six weeks later. Teachers report that the results add a layer of authority to the scores they give on classroom assessment. Subjecting the papers to an outside authority lent credence to their own evaluation of student work. From a system perspective, as our teachers analyze the externally scored work, they end up recalibrating themselves to the scoring guide.

At the beginning of the year there was some question as to whether the assessment was useful or was intrusive into teachers' instructional time. When the scored papers were returned, teachers made it clear that they wanted to repeat the assessment next fall. But the direct writing assessment is expensive. It costs more than seven dollars per student for scoring, and we question whether in tight times we can continue to support the assessment. Nonetheless, it is clear that when assessment can be used as part of the instructional program, teachers value the information it provides.

One major issue, then, is how we can support more instructionally useful assessment. It is clear that performance-based assessment needs to take place in the classroom and be evaluated by classroom teachers. One of the major implications for future testing is the need to train teachers to design assessments related to identified learning goals and to score them consistently (Stiggins, 2000). For this to happen will require coordinated ongoing efforts on the part of school districts, teachers' associations, principals' associations, state agencies, and institutions of higher education. Many experienced teachers believe that they know how to assess and are reluctant to discuss issues of consistency or even the degree to which their assessments measure the outcomes they really value.

On the other hand, technology may be able to assist with this type of assessment. Project Essay Grade (PEG), Intelligent Essay Assessor (IEA), and E-Rater are online tools designed to use artificial intelligence to score student writing (Rudner & Gagne, 2001). These tools are designed to provide instantaneous feedback to students and teachers on students' writing ability. Although some of these programs provide holistic scores, others provide more extensive trait-based scoring.

For now these systems require that students respond to preselected writing prompts, meaning that these kinds of programs are not open to teachers who want to assign students a content-based essay in literature, social studies, or science; given the rate of improvement in technology as we know it, however, it may not be that far in the future before electronic essay grading is generally available to teachers, allowing them to assign and grade more work. Even more exciting is the

possibility that students will be able to submit their essays for electronic scoring, get feedback, and revise their work until they achieve the appropriate level of accomplishment.

Similarly, with voice technology changing at a rapid rate, students may in the not-to-distant future be able to deliver a speech and receive feedback on the content and delivery—though not eye contact—instantaneously. This could improve the consistency and frequency of the evaluation of oral language skills, including speech, drama, and so forth.

But the impact of technology on assessment is not limited to the evaluation of essays or speeches. Technological solutions for testing include the following possibilities:

- online coursework with built-in assessment
- stand-alone systems designed to test one or two students at a time
- networked systems designed to test an entire class at the same time
- Internet-delivered tests, which students take online
- Internet-enabled tests, which students take via network-connected computers, where student data, test items, and scoring information are transmitted online from the testing organization (Olson, 2002)

The possibilities for delivering computerized adaptive tests to students are not new, but as the technological delivery systems become less expensive and easier to use they will make student assessment much more efficient. The benefit of adaptive testing is the ability to get information quickly about a student's level of performance. The testing can quickly focus on a student's achievement level and not waste time giving items that are too easy or too difficult for the student.

Each of these models promises teachers and students almost instantaneous score reporting and feedback. The models also make record-keeping easier by collecting student assessment data into databases that teachers, students, and even parents can potentially access.

So if classroom assessment becomes increasingly efficient, reliable, and standards-based, perhaps there will be less need for standardized district or state assessment. Alternatively, perhaps we will continue to need large-scale assessment to track our progress on our identified learning outcomes. If so, one of the issues we have to deal with is how much time we spend engaged in large-scale assessment. In

this district the amount of time spent in district or statewide assessment ranges from no time at 5th and 11th grades to 10 to 12 hours at 4th, 6th, 7th, and 10th grades. These times include only testing time, not the time spent getting students ready for testing or organizing the testing schedule, or the learning time lost, especially in secondary schools, where schedules of students not being tested are disrupted to accommodate the needs of those being tested. Although testing time is not lost time for those being tested, it can have a major impact on other instruction and it can be disruptive to the school program as a whole.

And what about accountability? ESEA calls for improvement of all groups, including all ethnic groups, students with disabilities, disadvantaged students, and students with limited English proficiency. With all the weight of accountability on the state assessments at grades three through eight, there will be major issues regarding test security, ethical test practice, and test inclusion. What about those students for whom our standardized testing is inappropriate? A major thrust in future assessment has to be identifying ways of incorporating results from alternate assessment into the main assessment system.

And then what about students who refuse to participate, or whose parents refuse to include them, in the testing system? Whether they object to the stakes of testing, the time taken by testing, or the limited sample that can be included in a large-scale assessment, parents have been organizing at the grassroots level to oppose large-scale testing and to boycott the tests. The impact has varied across the states, but it does represent a concern that could have major effects on the future of state and district programs.

The future of testing may be exciting—bringing high-quality data into the hands of teachers, students, and their parents so that all students can be well taught and will develop the skills and knowledge they need to be successful workers in the twenty-first century. New testing technology, increased teacher classroom assessment skills, and better record-keeping systems are all trends that will improve the quality of learning for students. On the other hand, heavy accountability requirements and testing that is limited to those constructs that are most easily assessed could have a devastating effect on learning and teaching.

As we wrap up the final boxes, making sure that we can account for each student's test booklets, every teacher's manual, all the alternate assessment forms for students who could not because of disability participate in the regular testing program, we imagine that next year's testing program will look similar to this year's program. But what will it look like in 5 or 10 years?

It is difficult to get outside the realities of today and try to project 5 to 10 years into the future, but why not try? It is 2012, and Mr. Harada's 9- and 10-year-old students are busy with reading and writing activities. Mr. Harada sits down with Yusif to listen as he reads aloud. The tablet Mr. Harada carries is connected via wireless network to the district computer system, and as Yusif picks up his reader, Mr. Harada gets the text of the passage on his tablet. He clicks "start" as he nods at Yusif to begin reading. As Yusif reads, Mr. Harada makes marks on his tablet. At the end he marks "finished" on his tablet and proceeds to discuss the passage with Yusif while making some notes on the child's comprehension. When they are done, Mr. Harada sends his notes to the computer and a report is generated that includes Yusif's reading and error rate, an analysis of reading errors, and a measure of his level of comprehension. With a tap of his stylus, Mr. Harada makes a report appear on his tablet, and he and Yusif discuss the results. Although he is still struggling some, Yusif has made great strides in reading. He identifies some areas that he needs to work on. Mr. Harada identifies some "next tasks" for Yusif and directs him to the classroom library to pull another book off the shelf. Yusif asks if he can take a copy of the report home to his grandmother, and with another tap of the stylus, Mr. Harada rolls a report off the back of his tablet. Mr. Harada can just as easily e-mail a copy of the report from his tablet, but today he knows that Yusif wants to hand the report to his grandmother himself. At any time, however, parents or guardians can check their child's work online and receive a complete report on the child's level of performance, and what next steps the child needs to take.

A few minutes later Mr. Harada sits down at his desktop computer and generates an analysis of his class' performance in reading skills. With this in hand, he takes a few minutes to decide which group of students he is going to pull together next for direct instruction. Between the oral reading assessments he conducts regularly and the assignments that students complete online or on their tablets, Mr. Harada is able to get a pretty complete picture of student progress toward meeting the district grade-level objectives.

Similarly, in the principal's office, or at the district office, an administrator can pull up a report that summarizes reading proficiency in Mr. Harada's class, or among students at the building or district level. Students still participate in the state-required assessments in reading and mathematics at grades three through eight, but these assessments are on their way out. The results of classroom-based assessment have been shown to correlate so highly with the state assessments that the

state has concluded that the large-scale assessment is redundant and an unnecessary expense. The money is better spent in training teachers in instruction and integrating assessment into instruction. With new tools coming on line at a rapid pace, all designed to make the assessment process effortless for teachers, keeping teachers and building administrators up to date is a major effort in all states.

At the national level, the National Assessment of Educational Progress (NAEP) is still administered using a matrix-sampling model, but it too has been shown to correlate highly with results reported by individual states based on classroom-level assessments. The president and members of Congress still believe that it is important to have an ongoing measure of student achievement, and NAEP satisfies their perceived need.

At the high school level, demonstration of mastery of essential skills is necessary for a high school diploma. Students begin to collect artifacts of their work at the beginning of ninth grade. As students transmit their work into their electronic portfolios, it is scored and retained. If work is judged not to be of sufficient merit at any time, it is returned to the students with feedback on where the project needs work. All along the way students have the opportunity to submit work and seek feedback, whether it is in math problem solving or reading and writing. At any point teachers can monitor student progress in gathering artifacts, and classes can be grouped or regrouped as needed to assist students in meeting various requirements.

Classrooms look fairly similar to those at the turn of the century, but what is different is the fact that students, being able to receive almost continuous feedback on their work, are motivated to succeed. Their understanding of scoring guides and expectations is simply built into the system, and they are typically able to judge the quality of their own work before either a scoring system or a teacher evaluates it. As teachers are almost totally freed from the drudgery aspect of evaluating student work, they can spend more time evaluating students' strengths and weaknesses, and plan and deliver appropriate instruction to large or small groups, depending on need.

Another major difference is that students are able to progress at their own rates. Although English classes still discuss literature, and foreign language classes continue to build oral communication, much of the skill building occurs within small groups of students. Ongoing assessment provides feedback to all, enabling teachers to focus instruction on those who need it.

The assessment community has not disappeared. Theoreticians are developing the tools to evaluate increasingly complex tasks. They are directing the work on artificial intelligence for evaluating student work, and continually testing the reliability and validity of the models. In addition, they have helped develop security systems that ensure that the student being assessed has actually done the work. Term-paper mills are a thing of the past, because their products won't pass the security system, and if Mom or Dad completes a student's work, that is obvious too. Students are used to the security systems, so it isn't really an issue—besides, the type of feedback they receive is so engaging that the concept of cheating is a foreign one.

Although each state and district has slightly different graduation requirements, students who move from one place to another can transfer their portfolios and have them assessed anew. The feedback they receive makes the process of updating their artifacts for a new system fairly easy.

Rewards for high-performing and punishment for low-performing schools are a thing of the past. Because schools can monitor student progress regularly, and systems are in place at the local and state levels to identify schools where significant numbers of students are not making progress, intervention can happen almost instantaneously. No child is left behind, because there are many resources available to track progress and intervene where needed. State and local SWAT teams can be directed to a site for a short time to work closely with classroom teachers. They can provide training and support for teachers, and provide direct intervention for students when needed. Because the assistance is short, and follow-up can be maintained, teachers welcome the support and assistance.

In 2012 assessment is almost totally embedded in instruction. Because teachers, administrators, and parents can easily monitor results, instruction can be tailored specifically to student needs. Formal large-scale assessments occur on occasion, primarily to ensure that the regularly gathered data are reliable and valid. The focus of teachers, administrators, parents, and students is on learning, and students have become key evaluators of their own achievement. From the outset, students know what they are expected to learn, how that learning will be evaluated, and what they need to do to get there. Teachers can focus on instruction, and administrators are instructional leaders, focused on continual training to help all teachers meet the needs of each of their students.

Back to 2003, the boxes are ready to go. And now we just have to wait patiently for three months until we can get some limited feedback on how well our students are meeting the standards.

References

Olson, A. (2002, April). Technology solutions for testing. *The School Administrator Web Edition.* Retrieved February 22, 2003, from http://www.aasa.org/publications/sa/2002_04/olson.htm.

Rudner, L., & Gagne, P. (2001). *An overview of three approaches to scoring written essays by computer.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. (Eric Document Reproduction Service No. ED458290)

Stiggins, R. (2000). *Student involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[ ] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[X] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").